Title: Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study

Author names: Tom White, Kate Westgate, Stefanie Hollidge, Michelle Venables, Patrick Olivier, Nick Wareham, Soren Brage

Author affiliations: MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom (TW, KW, SH, NW, SB)

MRC Elsie Widdowson Laboratory, 20 Fulbourn Road, Cambridge, United Kingdom (MV)

Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom (PO)

Corresponding Author: Soren Brage, MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Addenbrooke's Biomedical Campus, Cambridge CB2 0QQ, United Kingdom; +44 1223 769116; soren.brage@mrc-epid.cam.ac.uk

Sources of Support: Medical Research Council (http://www.mrc.ac.uk/) grants MC_UU_12015/1 and MC_UU_12015/3 to NW and SB, studentship from MedImmune to TW. Medical Research Council, UK Biobank, MedImmune and Newcastle University strategic funding for Digital Civics covered the costs of the field work.

Short running head: Energy expenditure from body-worn sensors

Abbreviations list: Activity Energy Expenditure (AEE), Doubly-Labelled Water (DLW), Diet-Induced Thermogenesis (DIT), Euclidean Norm Minus One (ENMO), Food Quotient (FQ), Food Frequency Questionnaire (FFQ), High-pass Filtered Vector Magnitude (HPFVM), Resting Energy Expenditure (REE), Total Energy Expenditure (TEE), Vector Magnitude (VM).

Abstract

Background: Many large studies have implemented wrist or thigh accelerometry to capture physical activity, but the accuracy of these measurements to infer Activity Energy Expenditure (AEE) and consequently Total Energy Expenditure (TEE) has not been demonstrated. The purpose of this study was to assess the validity of acceleration intensity at wrist and thigh sites as estimates of AEE and TEE under free-living conditions using a gold-standard criterion.

Methods: Measurements for 193 UK adults (105 men, 88 women, aged 40-66 years, BMI 20.4-36.6 $kg \cdot m^{-2}$) were collected with triaxial accelerometers worn on the dominant wrist, non-dominant wrist and thigh in free-living conditions for 9-14 days. In a subsample (50 men, 50 women) TEE was simultaneously assessed with doubly labelled water (DLW). AEE was estimated from non-dominant wrist using an established estimation model, and novel models were derived for dominant wrist and thigh in the non-DLW subsample. Agreement with both AEE and TEE from DLW was evaluated by mean bias, Root Mean Squared Error (RMSE) and Pearson correlation.

Results: Mean TEE and AEE derived from DLW were 11.6 (2.3) $MJ \cdot day^{-1}$ and 49.8 (16.3) $kJ \cdot day^{-1} \cdot kg^{-1}$. Dominant and non-dominant wrist acceleration were highly correlated in free-living (r=0.93), but less so with thigh (r=0.73 and 0.66, respectively). Estimates of AEE were 48.6 (11.8) $kJ \cdot day^{-1} \cdot kg^{-1}$ from dominant wrist, 48.6 (12.3) from non-dominant wrist, and 46.0 (10.1) from thigh; these agreed strongly with AEE (RMSE ~12.2 $kJ \cdot day^{-1} \cdot kg^{-1}$, r ~0.71) with small mean biases at the population level (~6%). Only the thigh estimate bias was statistically significantly different from the criterion. When combining these AEE estimates with estimated REE, agreement was stronger with the criterion (RMSE ~1.0 $MJ \cdot day^{-1}$, r ~0.90).

Conclusions: In UK adults, acceleration measured at either wrist or thigh can be used to estimate population levels of AEE and TEE in free-living conditions with high precision.

Keywords: physical activity; wrist acceleration; wrist-worn sensor; thigh-worn; isotope; bioenergetics; validation; energy balance

Introduction

Characterising the energy balance of individuals in free-living conditions requires an accurate assessment of total energy expenditure. Total energy expenditure can be measured with high precision using the doubly labelled water technique[1] but this is an expensive undertaking that requires elaborate sample collection and analysis infrastructure, making it less feasible for large-scale deployment or application in clinical settings. In most people, the largest component of total energy expenditure is resting energy expenditure, which can be predicted from anthropometric information with reasonable accuracy[2,3]. Diet-induced thermogenesis is less variable and ordinarily constitutes approximately 10% of total energy expenditure[4]. The predominant source of uncertainty in total energy expenditure estimates is the highly-variable activity energy expenditure component, which has proven difficult to capture by subjective instruments such as questionnaires[5,6]. Body-worn sensors such as accelerometers have the potential to provide a relatively cheap and reliable solution to this problem[7], if valid inference models can be devised to estimate activity energy expenditure from the measurements they record.

In recent years, wrist-worn accelerometers have become a popular measurement modality for objectively capturing free-living physical activity in large-scale studies[8–10]. Devices worn on the wrist are generally considered to be less burdensome for participants than those worn on other anatomical sites[11]. This has led to improved wear protocol adherence and thus to measurements with potentially greater representation of habitual physical activity levels. However, despite their recent increase in popularity, their utility in the estimation of activity energy expenditure has yet to be tested against gold-standard techniques in a sufficiently large sample of men and women in free-living[12]. Furthermore, some large studies [8–10] have committed to measuring only one of either the dominant wrist or non-dominant wrist, and the relationship between these two measurements also remains understudied.

In previous work, we derived parametric models to estimate activity energy expenditure intensity from non-dominant wrist acceleration (reproduced in Table 2) using a dataset (n=1050) of simultaneous non-dominant wrist and individually-calibrated combined heart rate and movement sensing signals collected under free-living conditions[13]. We evaluated the models in a large holdout sample (n=645) and found that they explained 44-47% of the variance in activity energy expenditure with no significant mean bias at the population level. However, as this comparison was against a silver-standard measurement of activity volume, these estimation models could be more conclusively validated by integrating the estimated activity energy expenditure signal over time, and assessing agreement of activity volume with a gold-standard criterion such as doubly labelled water. This approach has been used to validate combined heart rate and movement sensing [14–16] against which the models were originally derived.

Thigh-worn devices have typically been employed in smaller studies to measure time spent in a sitting posture, in order to infer sedentary time. This is possible because the distribution of gravity over the three axes can be interpreted using a simple equation to calculate thigh inclination. However, thigh acceleration has received comparatively little attention as a measure of physical activity intensity, though it features prominently in activity classification experiments[17]. In epidemiological settings, thigh-worn sensors have been complemented by other sensors with the intention to capture physical activity separately[18].

The primary aim of this study was to describe the absolute validity of a previously established activity energy expenditure prediction model [13] when applied to both wrists, and to evaluate the validity of this estimation in predicting total energy expenditure when combined with a simple anthropometric prediction of resting energy expenditure[2]. The second aim was to use the same approach to derive and validate similar energy expenditure estimation models using thigh acceleration. The third aim was to explore the relationship between the dominant wrist, non-dominant wrist and thigh acceleration measures in free-living, and to derive intensity models to facilitate harmonisation.

Subjects and Methods

Participants were recruited from the Fenland study, an ongoing cohort described in detail elsewhere [19]. We aimed to recruit participants who had previously indicated that they were interested in participating in future studies, were aged between 40 and 70 years, with a BMI between 20 and 50 $kg \cdot m^{-2}$. Recruitment aimed to balance age, sex and BMI distributions. Participants were invited to attend an assessment centre on two separate occasions, separated by a free-living period of 9 to 14 days. Ethical approval for the study was obtained from Cambridge University Human Biology Research Ethics Committee (Ref: HBREC/2015.16). All participants provided written informed consent.

Weight was measured to the nearest 0.1 kg using calibrated digital scales (TANITA model BC-418 MA; Tanita, Tokyo, Japan) at both visits. Height was measured to the nearest 0.1 cm using a stadiometer (SECA 240; Seca, Birmingham, UK) at the first clinic visit. Body composition was also measured by DXA (Lunar Prodigy Advanced, GE Healthcare, USA) as part of the Fenland study.

Total energy expenditure was measured by doubly labelled water in 100 of the participants. Prior to the first clinic visit, participants self-reported their current weight, which was used to provide a body-weight specific dose of $^{2}H_{2}{}^{18}O$ (70 mg $^{2}H_{2}O$ and 174 mg $H_{2}{}^{18}O$ per kg body weight). Participants brought a baseline urine sample to their first clinic visit, and a second baseline sample was taken at the clinic visit, prior to dosing. Participants were provided labelled sampling bottles and asked to collect one urine sample per day for the next 9-10 days, at a similar time each day but not the first void of the day. Participants were asked to record the date and time of each measurement on the sample bottle label and separately on a provided timesheet. Participants were asked to store the samples in a container in a cool, dry place, such as a refrigerator, and to return those samples at their second clinic visit at the end of their free-living measurement period. Isotope ratio mass spectrometry (^{2}H, Isoprime, GV

Instruments, Wythenshaw, Manchester, UK and ^{18}O, AP2003, Analytical Precision Ltd, Northwich, Cheshire, UK) was used to measure the isotopic enrichment of the samples. All samples were measured alongside laboratory reference standards, previously calibrated against the international standards Vienna-Standard Mean Ocean Water (vSMOW) and Vienna-Standard Light Antarctic Precipitate (vSLAP) (International Atomic Energy Agency, Vienna, Austria). Sample enrichments were corrected for interference according to Craig [20] and expressed relative to vSMOW. Rate constants and pool sizes were calculated from the slopes and intercepts of the log-transformed data, with total CO_2 production (RCO_2) calculated using the multi-point method of Schoeller [21]. RCO_2 was converted to total energy expenditure [22] where the respiratory quotient was informed by the macronutrient composition of the diet (see below).

Resting metabolic rate was measured at the start of both clinic visits during a fifteen-minute rest test in a supine posture by respired gas analysis (OxyconPro, Jaeger, Germany), after an overnight fast. Participants were asked not to eat or drink anything but water 2 hours prior to the appointment, and to refrain from smoking, chewing nicotine gum, wearing nicotine patches, or engaging in heavy physical activity. A seven-breath running median was calculated and the lowest observed average rate over a five minute consecutive window was found, which was scaled down by 6% to compensate for within-day elevation of resting metabolic rates [23]. Basal metabolic rate was also estimated via three different equations which differ in the specific body composition information utilised [2,24,25]. Resting energy expenditure was primarily characterised as the nearest measured value to the mean average estimated value, and a further sensitivity analysis was conducted using exclusively measured values. The final 24-hour resting energy expenditure estimates also included an adjustment for a 5% lower metabolic rate during sleep[26], according to their reported mean sleep duration.

At the second clinic visit, participants were asked to complete a Food Frequency

Questionnaire[27], which was used to estimate dietary intake over the past year. The food frequency data was processed using FETA[28], and the resulting calorie-weighted macronutrient profile was used to calculate the Food Quotient and diet-induced thermogenesis[29]. Diet-induced thermogenesis was normalised by the total energy expenditure to total energy intake ratio, as done previously[14].

At the first clinic visit, participants were fitted with three waterproof triaxial accelerometers (AX3, Axivity, Newcastle upon Tyne, UK); one device was attached to each wrist with a standard wristband, and one was attached to the anterior midline of the right thigh using a medical-grade adhesive dressing. The devices were setup to record raw, triaxial acceleration at 100 Hz with a dynamic range of ±8 g (where g refers to the local gravitational force, roughly equal to 9.81 $m \cdot s^{-2}$). Participants were asked to wear them continuously for the following 8 days and nights whilst continuing with their usual activities. They were also asked to record their main sleep using a sleep diary throughout the free-living period.

The signals were resampled from their original irregularly timestamped intervals to a uniform 100 Hertz signal by linear interpolation, and then calibrated to local gravity using a well-established technique[30,31], without adjustment for temperature changes within the record. Periods of nonwear were identified as windows of an hour or more wherein the device was inferred to be completely stationary [11], where stationary is defined as standard deviation in each axis not exceeding the approximate baseline noise of the device itself (10 milli-g). Vector Magnitude (VM) was then calculated from the three axes ($VM(X,Y,Z) = (X^2 + Y^2 + Z^2)^{0.5}$), from which two acceleration intensity metrics were derived [32]; Euclidean Norm Minus One (ENMO) subtracts 1 g from VM and truncates any negative results to 0, and High-Pass Filtered Vector Magnitude (HPFVM) applies a fourth-order high-pass filter to the signal at a 0.2 Hertz cut-off (3 dB). These analyses were performed using pampro v0.4.0[33].

In the non-doubly labelled water group (n=93), multi-level linear regression with random effects at the participant level was used to characterise each of the pairwise relationships between dominant wrist, non-dominant wrist and thigh acceleration intensity using synchronised 5-minute level data from each source. We used these intensity relationships to derive new activity energy expenditure estimation models for thigh and dominant wrist-worn devices, by substituting the non-dominant wrist term in our original models with the derived equation to harmonise either dominant wrist or thigh acceleration to non-dominant wrist acceleration.

Activity energy expenditure was estimated separately from each of the acceleration signals by directly applying the appropriate linear and quadratic equations given in Table 2 to 5-second level data; the resulting 5-second level estimated activity energy expenditure signal was then summarised to a mean-per-day average activity energy expenditure using diurnal adjustment to compensate for any between-individual bias introduced by periods of nonwear[34]. To ensure a stable estimate of this circadian model, a minimum of 72 hours of valid data was required per signal to be included in the analyses. Predicted total energy expenditure (in $\mathrm{MJ \cdot day^{-1}}$) was calculated as the sum of predicted activity energy expenditure and predicted resting energy expenditure from the simplest model (using only age, sex, height and weight)[2], and dividing the result by 0.9 to account for diet-induced thermogenesis[4]. Agreement between these two predictions against measured activity energy expenditure and total energy expenditure from doubly labelled water was formally tested by calculating the pairwise mean bias and 95% limits of agreement, Root Mean Squared Error (RMSE) and Pearson's correlation coefficient. Linear regression was used to characterise the relationship between the acceleration measurements and activity energy expenditure/total energy expenditure derived from doubly labelled water. As the main focus of this paper is on absolute validity, these relative validity results are supplied in the supplementary material.

The statistical tests were performed using Python v3.6 and Stata v14 (StataCorp, TX, USA).

Results

A descriptive summary of participant characteristics is given in Table 1. We recruited 193 participants, and the group measured by doubly labelled water was split equally between men and women. According to the doubly labelled water measurements, mean (standard deviation) total energy expenditure was 11.6 (2.3) MJ·day^{-1}, of which 6.6 (1.2) MJ·day^{-1} was resting energy expenditure. Mean (standard deviation) activity-related acceleration (ENMO) per day was 32.4 (8.3) milli-g on the dominant wrist, 28.8 (7.7) milli-g on the non-dominant wrist, and 27.8 (10.9) milli-g on the thigh. Mean dominant wrist acceleration was higher than non-dominant wrist in 84% of participants.

Some accelerometry measurements were not included in the analyses due to a combination of devices being lost by participants (n=7), device failures (n=3), user error upon download (n=3), and insufficient wear time (n=3). Of those files that overlapped with doubly labelled water measurements, 3 were dominant wrist records, 3 were non-dominant wrist and 9 were thigh records. There was no loss of data in the doubly labelled water, anthropometry or food frequency questionnaire measurements.

Table 2 lists the derived equations to predict activity energy expenditure from each of the sensors, as informed by the harmonisation equations which are supplied in Supplementary Table 1. For brevity, Table 3 summarises the absolute validity of the quadratic HPFVM models applied to measurements from both wrists and thigh with respect to activity energy expenditure, and Table 3 summarises agreement with total energy expenditure derived from doubly labelled water. A Bland-Altman plot illustrating the agreement of these estimates is supplied in Figure 1. A table summarising the remaining models is given in Supplementary Table 2.

The difference in performance between each estimation model was very minor; all activity energy expenditure estimates had small negative mean biases (underestimates) at the population level (average -2.8 kJ·day^{-1}·kg^{-1}) but of these only the thigh model biases were statistically significant. RMSEs for activity energy expenditure ranged from 11.9 to 13.5 kJ·day^{-1}·kg^{-1} (24 to 27% of the mean), and 1.0 to 1.2 MJ·day^{-1} for total energy expenditure (8 to 10% of the mean). Pearson correlations ranged from 0.6 to 0.69 with activity energy expenditure, and from 0.87 to 0.91 with total energy expenditure. Combined estimates using two or more sensors lead to very negligible performance improvements over single-sensor estimates. Signed estimation errors were nominally positively correlated with body fat percentage when using our primary characterisation of resting energy expenditure (r=0.18-0.25), and less so with exclusively measured values (r=0.10-0.17). For each estimate there was a significant trend of overestimation in the least-active to underestimation in the most active (mean trend r=0.7 for activity energy expenditure and 0.45 for total energy expenditure).

In the non-doubly labelled water group, 88 participants had at least 3 days of valid simultaneous wrist signals during free-living, and 84 had simultaneous wrist and thigh signals; around 200 000 5-minute observations were included in each of the regression analyses. The between-individual explained variance between dominant and non-dominant wrist intensity signals was approximately 86% (99% within-individual), and the average between-individual explained variance between wrist and thigh intensities was approximately 49% (97% within-individual). The derived linear models to harmonise the acceleration signals are listed in Supplementary Table 1. The final models given to estimate activity energy expenditure from dominant wrist and thigh in Table 2 were the result of substituting these harmonisation equations into the original non-dominant wrist models.

Discussion

In this work, we have applied our previously derived activity intensity estimation models [13] to wrist acceleration signals (after harmonising the intensity of dominant wrist to non-dominant wrist) and investigated their agreement with a gold-standard measure of activity energy expenditure. We arrived at estimates that were moderately correlated with the criterion ($r > 0.6$) with small and non-significant mean biases at the population level from both wrists and RMSEs of approximately 12 $kJ \cdot day^{-1} \cdot kg^{-1}$. We have also introduced and validated new intensity estimation models for thigh acceleration, demonstrating similar performance to the wrist models. We then used the activity energy expenditure estimates to model total energy expenditure by combining with anthropometry-based predictions of resting energy expenditure; we found stronger agreement with the criterion ($r=0.9$, RMSE=1.0 $MJ \cdot day^{-1}$) due in part to the relatively high accuracy of resting energy expenditure prediction equations. We observed that dominant wrist acceleration was on average 12% higher than non-dominant wrist in free-living individuals, but that those measures were very highly correlated ($r=0.93$), allowing us to derive conversion models which harmonise acceleration intensity measured at either wrist. To our knowledge, this is the first demonstration of the absolute validity of a time-integrated predictive model of activity intensity for either wrist or thigh accelerometry. Our findings on the high correlation between dominant wrist and non-dominant wrist acceleration in free-living individuals are consistent with a previous study in a small convenience sample (n=40)[35]. They also observed ~5% higher dominant wrist than non-dominant wrist acceleration, but it was not a statistically significant difference, perhaps due to the shorter duration of measurement and smaller sample size. In our relative validity tests, we found that each wrist separately explained a similar variance in activity energy expenditure, and inclusion of both wrist measurements in the linear models did not drastically improve performance over either wrist measurement alone. Taken together, these results are indicative of a high degree of upper-body symmetry. One implication of these findings is that irrespective of hand dominance, wrist acceleration measurements are naturally conducive to harmonisation across studies, making them well suited to pooled- and meta-analysis. Conversely, it implies that implementing dual wrist measurements may be a largely redundant exercise for studies whose primary intention is to capture activity energy expenditure. However, there is a possibility that future methodological advances in the field of activity recognition may be able to better utilise simultaneous wrist signals, which could yield a more precise instantaneous estimation of activity energy expenditure.

The estimation models validated herein for the wrist were derived using a training dataset in which non-dominant wrist acceleration data was collected at 60 Hz with a GeneActiv device [13], and were successfully validated using 100 Hz data collected with an Axivity AX3. The acceleration sampling frequency difference proved not to be an issue, because both likely satisfy the Nyquist sampling theorem across most or all human activities, and the models use mean movement intensity calculated over a 5-second window which make them robust to the number of samples that contribute to that mean.. With an additional harmonisation step, the model also translated to acceptably strong inferences on the dominant wrist, albeit with a slightly increased error. This indicates that our models capture a generalized biomechanical relationship of wrist movement, rather than being superficial transformations of a specific device's output to activity energy expenditure. It therefore suggests that these models are applicable to any wrist-worn device which provides raw, unfiltered triaxial acceleration data expressed in SI units.

The associations between wrist acceleration and observations from DLW have been reported before, in pregnant and non-pregnant Swedish women [11]. In that population it explained 27% of the variance in activity energy expenditure ($kJ \cdot day^{-1} \cdot kg^{-1}$) in non-pregnant women (n=48), but only 5% in pregnant women (n=26); however, those wrist measurements were evenly divided between left and right wrist, which most likely lead to a mix of dominant and non-dominant wrist measurements and potentially attenuated the correlations.

The previously established estimation models applied to the non-dominant wrist resulted in robust estimates with small, non-significant mean biases, which is a strong justification for using this inference scheme to infer activity energy expenditure in free-living individuals. The higher average of the dominant wrist would have led to a significant overestimation had we applied the original non-dominant wrist model, but our harmonisation approach effectively scaled the dominant wrist measure down to the level of non-dominant wrist, ultimately leading to virtually identical results. We used simple linear models to harmonise movement intensities between the different anatomical sites, which whilst evidently effective, may be improved upon in the future using more sophisticated techniques, such as nonlinear equations or neural networks. The Bland-Altman analyses showed trends of overestimation in the least active to underestimation in the most active across all estimation models, indicating that the models performed less precisely in absolute terms towards the extremes of high and low activity levels. These trends were stronger in the dominant wrist and thigh-based estimates, which may be a consequence of the additional harmonisation step causing an attenuation of the relationship.

We note that physical activity was measured by dominant wrist accelerometry in UK Biobank[8]. We have now demonstrated the validity of this approach in a demographically comparable sample. Specifically, the absolute validity result for ENMO in Supplementary Table 2 demonstrates that our linear estimation model applied to ENMO at 5-second resolution yielded a valid activity energy expenditure estimate, with a small mean bias and a RMSE of 13 $kJ \cdot day^{-1} \cdot kg^{-1}$ and moderately high correlation (r=0.61). Consequently, we can use the equations for dominant wrist in Table 2 to solve for salient energy expenditure values – for example, 3 metabolic equivalents (activity energy expenditure ~142 $J \cdot min^{-1} \cdot kg^{-1}$) is the generally accepted threshold for "moderate" activity intensity, and our ENMO equations suggest this is approximately 159 milli-g on the dominant wrist.

Our findings for the thigh acceleration models demonstrate that thigh-worn accelerometers capture an information-rich biomechanical signal, from which valid estimates of activity energy expenditure can be made. As a consequence of the larger y-intercepts of the thigh models, their minimum estimated activity energy expenditure ranges from 10 to 18 $J \cdot min^{-1} \cdot kg^{-1}$ (0.15-0.25 metabolic equivalents). To our knowledge, only one previous study has described the association between thigh acceleration and activity energy expenditure from doubly labelled water, in a small study of free-living cancer patients and controls[36]; which reported very low agreement between the manufacturer's proprietary activity energy expenditure prediction and the criterion. While thigh-worn sensors do not yet have the same popularity as wrist-worn sensors[37,38], large-scale data collections are planned for the future[39]. Our models enable new analyses to be conducted in those existing datasets, and may make thigh-worn accelerometry a more appealing option for future studies if issues of feasibility can be addressed.

Some have suggested that simple movement intensity approaches should be replaced by more sophisticated models that utilise a broader range of signal features[40,41]. Recent efforts to estimate energy expenditure have utilised a range of machine learning approaches, such as neural networks [42–44] and random forests[40]. While we are not aware of any such methodology with a performance that exceeds the simpler models validated in this paper, this is an interesting area of future work.

The results of our absolute validity tests demonstrate that deriving intensity models using a "silver-standard" criterion (such as individually-calibrated heart rate and uniaxial movement sensing) in a large sample of free-living adults is a sound approach. The combined sensing estimate of activity energy expenditure is less precise than respiratory gas analysis which can be captured in laboratory studies [45] but there are several reasons why we have been able to derive superior models to previous approaches. Firstly, the dataset was collected in free-living participants, and is therefore representative of the intended application, as opposed to artificial scenarios and activities performed in a laboratory. Secondly, the combined sensing approach embedded in a cohort study allowed the collection of a volume of data many orders of magnitude greater than any laboratory study has for this purpose. Our training dataset alone contained over 16.6 person-years of observation (>1.7 million data points). One disadvantage of this approach is that we are unable to capture categorical labelled data, so there is no opportunity to explore activity type recognition.

It is appropriate to compare our absolute validity results here with those of combined sensing itself [14]. The best estimate with treadmill test calibration resulted in a RMSE of 20 $kJ \cdot day^{-1} \cdot kg^{-1}$ (30% of the 66 $kJ \cdot day^{-1} \cdot kg^{-1}$ criterion mean), non-significant positive mean bias of approximately 4 $kJ \cdot day^{-1} \cdot kg^{-1}$ (6%) at the population level, and a correlation of 0.67 in a sample of 50 UK adults. Compared to the present results, all estimations here had considerably lower RMSEs of around 12 $kJ \cdot day^{-1} \cdot kg^{-1}$ (25% of the 50 $kJ \cdot day^{-1} \cdot kg^{-1}$ mean), similar magnitude but negative mean biases (~6%), but generally higher correlations. However, our study participants were significantly less active overall according to the criterion, ultimately leading to a similar relative accuracy. Combined sensing model errors were also uncorrelated to body fat percentage, whereas errors of accelerometry-only models seem to display this characteristic, albeit less so in the present study ($r=0.22$ versus $r=0.63$ for uniaxial trunk acceleration). Contrasting the feasibility of the methods, however, wrist accelerometry has the advantages of being cheaper, less burdensome to both participants and research staff, and does not require individual calibration using an exercise test. Comparing performance of other devices worn on the upper limbs, validation of the now-discontinued SenseWear Pro3 and Mini also achieved no significant bias with respect to total energy expenditure, but with lower correlations (r=0.84) than any of our total energy expenditure models (r=0.9) and wider limits of agreement [46] and with lower feasibility. An evaluation of activity energy expenditure estimates based on waist-worn accelerometry in 683 adults observed a mean estimation bias of -2.5 $kJ \cdot day^{-1} \cdot kg^{-1}$ and 95% limits of agreement between -33 and 30 $kJ \cdot day^{-1} \cdot kg^{-1}$ [47]. Unlike our study design their measurements were not strictly simultaneous, so their results describe the ability of estimates to characterise the latent activity level of the population, for which uncertainty would be expected to be higher.

In summary, we have evaluated the absolute validity of intensity models of activity energy expenditure from wrist and thigh accelerometry, and concluded that they provide sufficiently precise and accurate estimates in free-living adults. With the addition of predicted resting energy expenditure to produce total energy expenditure, we found even stronger validity at the population level. Considering its feasibility, wrist accelerometry emerges as a viable candidate for deployment in a large scale studies, including physical activity surveillance and the prediction of total energy expenditure in dietary surveys.

Acknowledgements

We are very grateful to the participants who took part in this study. We thank the principal investigators of the Fenland study for allowing us to recruit from this study population, and the functional teams of the MRC Epidemiology Unit (Study Coordination, Field Epidemiology, Anthropometry, Data Management, IT) for supporting the study. We would like to specifically acknowledge Lewis Griffiths, Katie Palmer, and Eoin McNamara for their assistance in the data collection for this study, and Annie Schiff, Richard Salisbury and Nicola Kimber for study co-ordination and recruitment.

We would like to thank Eirini Trichia from the MRC Epidemiology Unit for processing the FFQ data with the FETA package. We would also like to thank the stable isotope team from the MRC Elsie Widdowson Laboratory: Priya Singh, Elise Orford and Kevin Donkers for the DLW preparation and analysis.

Medical Research Council and UK Biobank are acknowledged for covering costs of the fieldwork. Newcastle University and MedImmune are acknowledged for covering the costs of the doubly labelled water measurements.

Competing interests

Patrick Olivier was a founding director of Axivity Ltd. (2011-2014); his spouse is currently CEO and a director of Axivity (from 2014). The remaining authors declare no conflict of interest.

References

1 Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jequier E *et al.* Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *Am J Physiol - Regul Integr Comp Physiol* 1986; **250**: R823–R830.

2 Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. *Public Heal Nutr* 2005; **8**: 1133–1152.

3 Sabounchi NS, Rahmandad H, Ammerman A. Best-fitting prediction equations for basal metabolic rate: informing obesity interventions in diverse populations. *Int J Obes* 2013; **37**: 1364–1370.

4 Westerterp KR, Jonge L de, Bray G, Granata G, Brandon L, Weststrate J *et al.* Diet induced thermogenesis. *Nutr Metab (Lond)* 2004; **1**: 5.

5 Maddison, Ralph, Mhurchu CN, Jiang Y, Hoorn S Vander, Rodgers A *et al.* International Physical Activity Questionnaire (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): A doubly labelled water validation. *Int J Behav Nutr Phys Act* 2007; **4**: 61–71.

6 Golubic R, May AM, Benjaminsen Borch K, Overvad K, Charles MA, Diaz MJT *et al.* Validity of electronically administered Recent Physical Activity Questionnaire (RPAQ) in ten European countries. *PLoS One* 2014; **9**. doi:10.1371/journal.pone.0092829.

7 Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: New insights and validation studies. *Obes Rev* 2013; **14**: 451–462.

8 Doherty A, Jackson D, Hammerla N, Ploetz T, Granat MH, White T *et al.* Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers : The UK Biobank Study. *PLoS One* 2017; : 1–14.

9 da Silva ICM, van Hees VT, Ramires V V, Knuth AG, Bielemann RM, Ekelund U *et al.* Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. *Int J Epidemiol* 2014; **43**: 1959–1968.

10 Troiano R, Mc Clain J. Objective measures of physical activity, sleep, and strength in US National Health and Nutrition Examination Survey (NHANES) 2011--2014. In: *Proceedings of the 8th International Conference on Diet and Activity Methods*. 2012.

11 van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY *et al.* Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS One* 2011; **6**: e22922.

12 Plasqui G, Bonomi Ag, Westerterp KR. Daily physical activity assessment with accelerometers: new insights and validation studies. *Obes Rev* 2013; **14**: 451–462.

13 White T, Westgate K, Wareham NJ, Brage S. Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS One* 2016; **11**. doi:10.1371/journal.pone.0167472.

14 Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U *et al.* Estimation of Free-Living Energy Expenditure by Heart Rate and Movement Sensing: A Doubly-Labelled Water Study. *PLoS One* 2015; **10**: e0137206.

15 Villars C, Bergouignan A, Dugas J, Antoun E, Schoeller DA, Roth H *et al.* Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. *J Appl Physiol* 2012; **113**: 1763–1771.

16 Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC, Wareham NJ. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. *Int J Epidemiol* 2011; **40**: 112–20.

17 Bao L, Intille SS. Activity Recognition from User-Annotated Acceleration Data. 2004; : 1–17.

18 Healy GN, Wijndaele K, Dunstan DW, Shaw JE, Salmon J, Zimmet PZ *et al.* Objectively measured sedentary time, physical activity, and metabolic risk the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). *Diabetes Care* 2008; **31**: 369–371.

19 O'Connor L, Brage S, Griffin SJ, Wareham NJ, Forouhi NG. The cross-sectional association between snacking behaviour and measures of adiposity: the Fenland Study, UK. *Br J Nutr* 2015; **114**: 1286–1293.

20 Craig H. Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochim Cosmochim Acta* 1957; **12**: 133–149.

21 Schoeller DA. Recent Advances in Nutritional Sciences Recent Advances from Application of Doubly Labeled Water to Measurement of Human Energy. 1998; : 1765–1768.

22 Elia, M and Livesey G. Theory and validity of indirect calorimetry during net lipid synthesis. *Am J Clin Nutr* 1988; **47**: 591--607.

23 Haugen H a, Melanson EL, Tran ZV, Kearney JT, Hill JO. Variability of measured resting metabolic rate. *Am J Clin Nutr* 2003; : 1141–1144.

24 Nielsen S, Hensrud DD, Romanski S, Levine J a, Burguera B, Jensen MD. Body composition and resting energy expenditure in humans: role of fat, fat-free mass and extracellular fluid. *Int J Obes Relat Metab Disord* 2000; **24**: 1153–1157.

25 Watson LPE, Raymond-Barker P, Moran C, Schoenmakers N, Mitchell C, Bluck L *et al.* An approach to quantifying abnormalities in energy expenditure and lean mass in metabolic disease. *Eur J Clin Nutr* 2014; **68**: 234–240.

26 Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C *et al.* 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011; **43**: 1575—1581.

27 Bingham S, Gill C, Welch A, Cassidy A, Runswick S, Oakes S *et al.* Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* 1997; **26 Suppl 1**: S137–S151.

28 Mulligan AA, Luben RN, Bhaniani A, Parry-Smith DJ, O'Connor L, Khawaja AP *et al.* A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open* 2014; **4**: e004503.

29 Jequier E. Pathways to obesity. *Int J Obes* 2002; **26**: S18–S27.

30 van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM *et al.* Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol* 2014; **117**: 738–44.

31 Lukowicz P, Junker H, Tröster G. Automatic Calibration of Body Worn Acceleration Sensors. In: *International Conference on Pervasive Computing*. Springer Berlin Heidelberg, 2004, pp 176–181.

32 van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S *et al.* Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One* 2013; **8**: e61691.

33 White T. pampro v0.4.0. 2018. doi:http://doi.org/10.5281/zenodo.1187043.

34 Brage S, Westgate K, Wijndaele K, Godinho J, Griffin S, Wareham N. Evaluation of a method for minimising diurnal information bias in objective sensor data. In: *Int Conf Amb Mon Phys Act Mov*. 2013.

35 Dieu O, Mikulovic J, Fardy PS, Bui-Xuan G, Béghin L, Vanhelst J. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. *Clin Physiol Funct Imaging* 2017; **37**: 525-529.

36 Skipworth RJE, Stene GB, Dahele M, Hendry PO, Small AC, Blum D *et al.* Patient-focused endpoints in advanced cancer: Criterion-based validation of accelerometer-based activity monitoring. *Clin Nutr* 2011; **30**: 812–821.

37 Schram MT, Sep SJS, Van Der Kallen CJ, Dagnelie PC, Koster A, Schaper N *et al.* The Maastricht Study: An extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *Eur J Epidemiol* 2014; **29**: 439–451.

38 van der Berg JD, Stehouwer CDA, Bosma H, van der Velde JHPM, Willems PJB, Savelberg HHCM *et al.* Associations of total amount and patterns of sedentary

behaviour with type 2 diabetes and the metabolic syndrome: The Maastricht Study. *Diabetologia* 2016; **59**: 709–718.

39 Hessen H-O, Tessem AJ. *Human Activity Recognition with two Body-Worn Accelerometer Sensors*. 2016.http://hdl.handle.net/11250/2407624.

40 Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physio* 2015; **35**: 2191–2203.

41 Montoye AHK, Begum M, Zachary H, Pfeiffer KA. Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiol Meas* 2017; **38**. doi:10.1088/1361-6579/38/2/343.

42 Montoye AHK, Mudd LM, Biswas S, Pfeiffer KA. Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free-Living. *Med Sci Sport Exerc* 2014. doi:10.1249/MSS.0000000000000597.

43 Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. *Med Sci Sports Exerc* 2014; **46**: 386.

44 Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiol Meas* 2016; **37**: 1770–1784.

45 Brage S, Brage N, Franks PW, Ekelund U, Wong M-Y, Andersen LB *et al.* Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *J Appl Physiol* 2004; **96**: 343–351.

46 Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Med Sci Sports Exerc* 2010; **42**: 2134–2140.

47. Matthews CE, Keadle SK, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, et al. Measurement of Active and Sedentary Behavior in Context of Large Epidemiologic Studies. Med Sci Sport Exerc [Internet]. 2017;(6):1. Available from: http://insights.ovid.com/crossref?an=00005768-900000000-97112

Figure legends

Table 1: Participant characteristics, provided separately for the doubly labelled water and non-doubly labelled water groups.

Table 2: Derived linear and quadratic equations to estimate activity energy expenditure ($J \cdot min^{-1} \cdot kg^{-1}$) from wrist and thigh acceleration intensity. (4.184 $J \cdot min^{-1} \cdot kg^{-1}$ = 1 cal, and 71.225 $J \cdot min^{-1} \cdot kg^{-1}$ = 1 net Metabolic Equivalent Task (MET)).

Table 3: Agreement between estimated activity energy expenditure from the HPFVM quadratic models with those derived from doubly labelled water. Bias values in bold indicate statistical significance according to a paired t-test ($p < 0.05$).

Figure 1: Bland-Altman plots illustrating agreement between the activity energy expenditure and total energy expenditure estimates from HPFVM Quadratic models with those from doubly labelled water, where the X-axis indicates the observed values.

Supplemental Table 1: Harmonisation equations relating movement intensities between the dominant wrist, non-dominant wrist and thigh.

Supplemental Table 2: Agreement between estimated activity energy expenditure from the all models with those derived from doubly labelled water. Bias values in bold indicate statistical significance according to a paired t-test ($p < 0.05$).

Supplemental Table 3: Derived regression models of activity energy expenditure (normalised for body weight) using all combinations of dominant wrist, non-dominant wrist and thigh acceleration.

Supplemental Table 4: Derived regression models of activity energy expenditure (not normalised for body weight) using all combinations of dominant wrist, non-dominant wrist and thigh acceleration, including body weight.

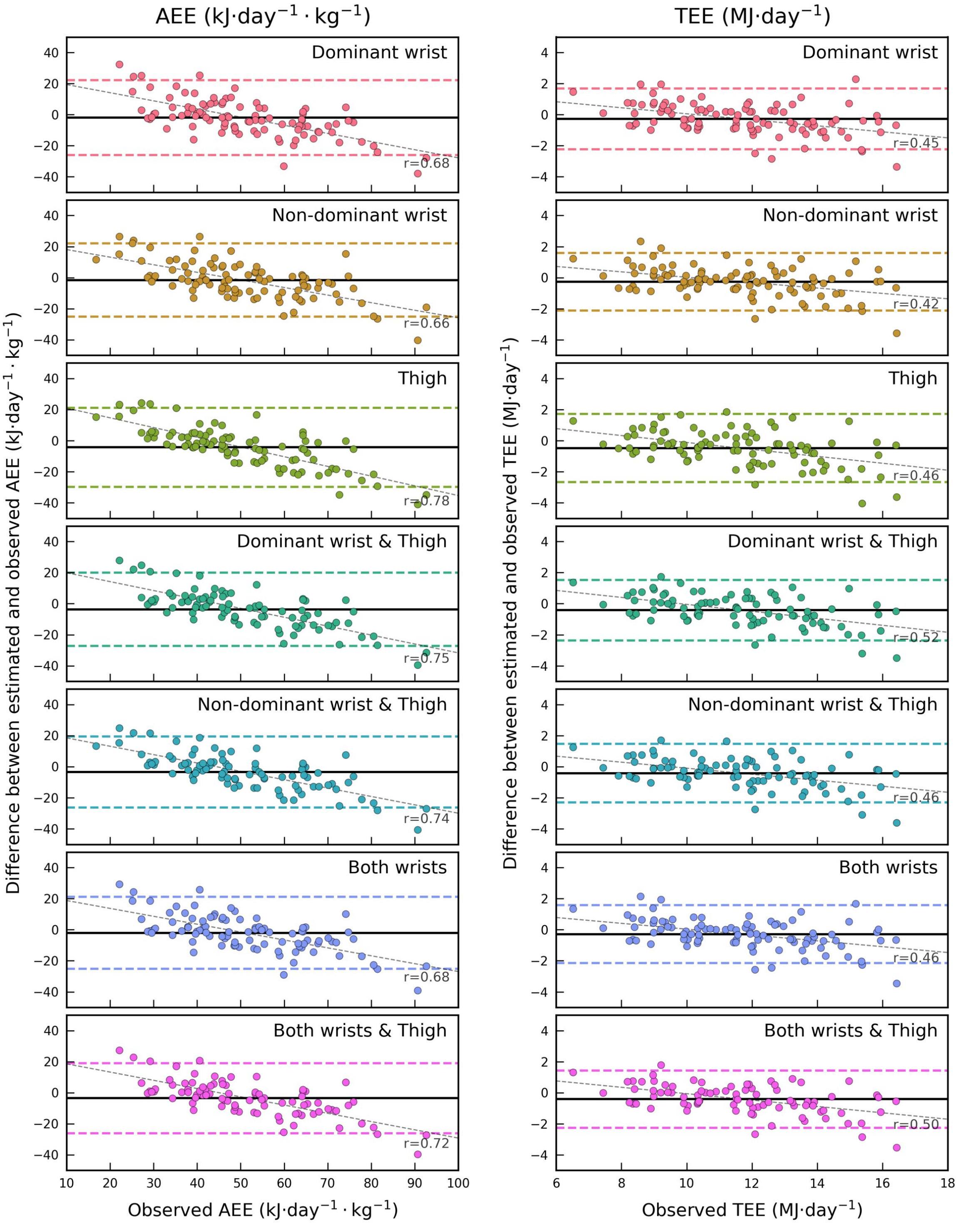
Supplemental Table 5: Agreement between estimated activity energy expenditure from the HPFVM quadratic models with those derived from doubly labelled water, in only right-handed individuals.

Supplemental Figure 1: Bland-Altman plots illustrating agreement between the activity energy expenditure and total energy expenditure estimates from ENMO linear models with those from doubly labelled water, where the X-axis indicates the observed values.

Supplemental Figure 2: Bland-Altman plots illustrating agreement between the activity energy expenditure and total energy expenditure estimates from ENMO quadratic models with those from doubly labelled water, where the X-axis indicates the observed values.

Supplemental Figure 3: Bland-Altman plots illustrating agreement between the activity energy expenditure and total energy expenditure estimates from HPFVM linear models with those from doubly labelled water, where the X-axis indicates the observed values.

Supplemental Figure 4: Bland-Altman plot illustrating the agreement between estimated resting energy expenditure using anthropometric equations and measured resting energy expenditure during the clinic visits.



AEE (kJ·day−1 · kg−1)
TEE (MJ·day−1)
Dominant wrist
Non-dominant wrist
Thigh
Dominant wrist & Thigh
Non-dominant wrist & Thigh
Both wrists
Both wrists & Thigh
r=0.68
r=0.66
r=0.78
r=0.75
r=0.74
r=0.68
r=0.72
r=0.45
r=0.42
r=0.46
r=0.52
r=0.46
r=0.46
r=0.50
Difference between estimated and observed AEE (kJ·day−1 · kg−1)
Difference between estimated and observed TEE (MJ·day−1)
Observed AEE (kJ·day−1 · kg−1)
Observed TEE (MJ·day−1)

	DLW (n=100)				Non-DL	
	Mean	Std.Dev.	Min	Max	Mean	Std.Dev.
Sex (% women)		50%				41
Age (years)	54.4	7.2	40.0	65.0	54.0	6.7
Height (m)	1.71	0.09	1.51	1.94	1.72	0.10
Weight (kg)	78.2	13.6	48.7	110.8	77.1	12.4
BMI (kg/m^2)	26.5	3.4	20.4	36.6	25.9	2.9
TEE (MJ/day)	11.60	2.32	6.52	16.43	-	-
REE (MJ/day)	6.61	1.24	3.74	9.86	-	-
DIT fraction	0.10	0.01	0.08	0.12	-	-
AEE (MJ/day)	3.87	1.38	0.72	7.56	-	-
AEE (kJ/day/kg)	49.8	16.3	8.5	92.6	-	-
k_O	0.119	0.03	0.066	0.257	-	-
k_H	0.093	0.028	0.044	0.228	-	-
N_O (moles)	2124	434	1215	3131	-	-
N_H (moles)	2188	447	1251	3224	-	-
DW ENMO (mg)	32.4	8.3	15.4	64.7	33.1	10.5
NDW ENMO (mg)	28.8	7.7	15.6	59.0	29.3	8.3
Thigh ENMO (mg)	27.8	10.9	13.2	76.3	28.2	10.0
DW HPFVM (mg)	48.5	11.0	25.7	85.9	49.6	12.8
NDW HPFVM (mg)	43.5	10.3	25.8	85.4	44.7	11.0
Thigh HPFVM (mg)	37.4	12.7	17.7	77.0	38.6	11.8

N (n=93)	
Min	Max
%	
41.0	66.0
1.53	1.96
56.4	112.3
20.4	35.3
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
18.8	82.4
16.2	63.2
12.6	80.5
31.4	105.7
27.3	89.2
17.7	94.6

Placement	Metric
NDW*	ENMO
NDW*	ENMO
NDW*	HPFVM
NDW*	HPFVM
DW	ENMO
DW	ENMO
DW	HPFVM
DW	HPFVM
Thigh	ENMO
Thigh	ENMO
Thigh	HPFVM
Thigh	HPFVM

(*) Published in White e

(**) x refers to accelera

Formulae to estimate AEE in J/min/kg (**)
5.01 + 1.000*x
-10.58 + 1.1176*x + 2.9418*sqrt(x) - 0.00059277*(x^2)
-4.65 + 0.8537*x
-1.25 + 1.1353*x - 2.4281*sqrt(x) - 0.00040270*(x^2)
5.01 + 1.000*(1.5 + .8517*x)
-10.58 + 1.1176*(1.5 + .8517*x) + 2.9418*sqrt((1.5 + .8517*x)) - 0.00059277*((1.5 + .8517*x)^2)
-4.65 + 0.8537*(1.3 + .8781*x)
-1.25 + 1.1353*(1.3 + .8781*x) - 2.4281*sqrt((1.3 + .8781*x)) - 0.00040270*((1.3 + .8781*x)^2)
5.01 + 1.000*(13.4 + .5674*x)
-10.58 + 1.1176*(13.4 + .5674*x) + 2.9418*sqrt((13.4 + .5674*x)) - 0.00059277*((13.4 + .5674*x)^2)
-4.65 + .8537*(20.3 + .6401*x)
-1.25 + 1.1353*(20.3 + .6401*x) - 2.4281*sqrt((20.3 + .6401*x)) - 0.00040270*((20.3 + .6401*x)^2)

et al, PLoS One 2016

ation (milli-g) measured at the relevant anatomical site, characterised with the relevant metric

Placement	Activity energy expenditure (kJ/day/kg)						Total en
	N	Bias (*)	95% LoA		r	RMSE	N	Bias (*)
Dominant wrist	97	-1.9	-26.0	22.2	0.644	12.4	97	-0.3
Non-dominant wrist	97	-1.5	-25.1	22.1	0.676	12.1	97	-0.3
Thigh	91	**-4.2**	-29.6	21.2	0.599	13.6	91	-0.5
Both wrists	94	-1.9	-25.1	21.3	0.669	11.9	94	-0.3
Non-dominant wrist & Thigh	89	-3.3	-26.2	19.6	0.687	12.1	89	-0.4
Dominant wrist & Thigh	88	-3.5	-27.2	20.1	0.644	12.5	88	-0.4
Both wrists & Thigh	86	-3.4	-25.9	19.2	0.675	11.9	86	-0.4

(*) Bias estimates in bold are statistically significant at p<0.05. (None of the TEE estimates were statistically s

ergy expenditure (MJ/day)			
95% LoA		r	RMSE
-2.2	1.7	0.903	1.0
-2.1	1.6	0.911	1.0
-2.7	1.7	0.874	1.2
-2.1	1.6	0.911	1.0
-2.3	1.5	0.909	1.0
-2.4	1.5	0.902	1.1
-2.2	1.4	0.914	1.0

significantly different.)